

La disinformazione come minaccia per la democrazia in Italia. Crisi della fiducia e rischi della società digitale

Olimpia Affuso, *Università degli Studi della Calabria*

Antonella Coco, *Centro Alti Studi Difesa - Scuola Superiore Universitaria*

This article has been accepted for publication, but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record.

Please cite this article as:

Affuso O., Coco A. (2026), *La disinformazione come minaccia per la democrazia in Italia. Crisi della fiducia e rischi della società digitale*, in «Cambio. Rivista sulle trasformazioni sociali», doi: 10.36253/cambio-19242.

La disinformazione come minaccia per la democrazia in Italia. Crisi della fiducia e rischi della società digitale

OLIMPIA AFFUSO

Università degli Studi della Calabria

ANTONELLA COCO

Centro Alti Studi Difesa - Scuola Superiore Universitaria

Contact: olimpia.affuso@unical.it

Abstract. This contribution analyses disinformation as a factor of vulnerability in contemporary democracies, focusing on the relationship between informational disorder, the erosion of trust and the weakening of democratic cohesion. It examines disinformation activities conducted by foreign state actors against Italy and investigates how institutional actors involved in monitoring and countering the phenomenon interpret and problematize them. The article argues that transformations in the digital media ecosystem reshape the cognitive and symbolic mechanisms through which trust, shared orientations and democratic legitimacy are produced.

The analysis combines scientific literature, national and supranational documentary sources, and qualitative evidence from semi-structured interviews with key informants in politics, defence, communication and research. The findings identify recurring dimensions: the framing of disinformation as a systemic, permanent and difficult-to-attribute threat; its effects on the credibility of information sources and institutional legitimacy; the governance difficulties it raises; and the preference for non-securitarian responses based on social resilience, critical capacities and cognitive security. The article also discusses the ambivalent role of artificial intelligence, both as an amplifier of disinformation and as a possible countermeasure, while stressing the limits of purely technological solutions. Overall, informational disorder emerges not only as the circulation of false content, but as a broader condition of informational hypertrophy that opens multiple access points for strategic manipulation and deepens confusion in public opinion.

Keywords: Disinformation; Hybrid conflicts; Institutional trust; Artificial intelligence; Democracy

INTRODUZIONE

L'impiego di strategie comunicative volte a gestire e controllare in vario modo l'informazione rappresenta da sempre una risorsa decisiva nelle relazioni di potere. Questa dimensione informazionale richiede importante attenzione e una riflessione urgente perché vi si generano momenti di "disordine informativo" via via più frequenti e la cui portata e criticità si sta ampliando notevolmente, assumendo una grande rilevanza non solo nel dibattito pubblico ma come problema da affrontare in termini di *policy* (Lovari, Ducci 2022).

Tale "disordine informativo" è l'effetto di una circolazione mista, "impastata" di informazioni e notizie distorte, false, ambigue, infondate, manipolatorie, e di informazioni basate, invece, su fonti accertabili e accreditate, affidabili. Rispetto a questo insieme, nel *Code of Practice on Disinformation*, la Commissione Europea definisce "disinformazione" il sottoinsieme che comprende informazioni e

notizie false o fuorvianti che «sono create, presentate e diffuse per un guadagno economico o per ingannare intenzionalmente il pubblico; possono causare danni pubblici, intesi come minacce ai processi democratici politici e decisionali, nonché a beni pubblici come la protezione della salute dei cittadini dell'UE, dell'ambiente o della sicurezza». Non si può tuttavia sottovalutare il ruolo che hanno anche altri tipi di notizie, come quelle che vengono messe in circolazione senza adeguata attenzione alle fonti, o condivise in buona fede senza intento di nuocere.

Sebbene incidano a vari gradi di criticità, si possono individuare diversi tipi di notizie che concorrono al disordine informativo. Si hanno *notizie confuse* quando nell'orientamento all'oggettività prevale un eccesso di soggettività, a causa del coinvolgimento di individui singoli e "disintermediati"; *notizie errate* quando nascono da incompetenza o disattenzione professionale. Si produce in questi casi *misinformazione*, in quanto non c'è intento di causare un danno. Si parla invece di *disinformazione* strettamente intesa di fronte a notizie false create volutamente per danneggiare avversari politici, concorrenti di mercato, o altri Paesi, come può avvenire durante un conflitto, come precisato dalla Commissione Europea. Vi rientrano almeno tre tipi di notizie: *menzognere*, capaci di alterare la realtà attraverso la falsificazione dei dati e delle conoscenze; *ingannevoli*, orientate a indurre nell'errore; *deformate*, che esagerano la realtà, come nel caso del giornalismo sensazionalistico, o di una informazione eccessivamente di parte costruita da giornalisti «subalterni» al potere politico (Carlson 2009, Boccia Artieri, Colombo & Gili 2022). In tutti e tre i casi si produce una manipolazione informativa intenzionale. A tali tipi si aggiungono le *notizie ostili*, o discorsi d'odio, che mirano a distruggere la reputazione di individui e organizzazioni. Si tratta invece di *malinformazione* quando il messaggio è autentico, ma viene condiviso per danneggiare una persona – come quando vengono portati in pubblico fatti privati di qualcuno. Il dato problematico è che tutti questi tipi di notizie riescono a fornire interpretazioni, spesso anche soluzioni a problemi, alterando la conoscenza dei fatti e la stessa percezione della realtà in modi che alla fine risultano inattesi, ingestibili, spesso pericolosi.

L'insieme di tali notizie, che alcuni hanno definito «cattiva informazione» (Giglietto *et alii* 2019; Jack 2017), produce disorientamento, scetticismo, diffidenza proprio in relazione al fatto che i confini tra informazione e disinformazione diventano indistinguibili. *Questa con-fusione* tra notizie attendibili e notizie false o distorte dà luogo a una nuova condizione informazionale, che assume i tratti di un *caleidoscopio*, in cui l'informazione, in maniera *mimetica*, si diffonde con infinite connotazioni in virtù delle emozioni, paure, *bias* cui aderisce, senza dover rispondere a un criterio di verità dei fatti, oppure trattenendo pezzi di verità per assumerne lo statuto. In questo sistema la manipolazione dei messaggi avviene in varie modalità, tramite video, meme, titolo clickbait; l'autore del contenuto può essere un troll pagato, o un professionista che ha commesso un errore, o che ha una strategia; il pubblico può avere competenze per riconoscere la distorsione o esserne incapace.

L'intero processo può essere ricondotto ad importanti cambiamenti e fattori, sia sul piano politico-culturale sia sul piano dell'innovazione tecnologica, alcuni dei quali intervenuti già da tempo, come la crisi delle narrazioni e delle ideologie del Novecento, altri più recenti, come l'avvento dei social media e dei relativi processi di disintermediazione informativa, ai quali fa da corollario un sempre più radicale indebolimento della fiducia in alcune istituzioni politiche e in alcuni sistemi esperti, compreso quello dell'informazione (Edelman 2022)¹. Il verificarsi di eventi particolari, come la pandemia e le recenti guerre, hanno reso ancora più visibile il problema, richiamando l'attenzione da parte delle istituzioni pubbliche e delle autorità governative nazionali e internazionali sul rischio che la disinformazione comporta per i sistemi sociali e per i cittadini.

Rispetto a tutto ciò, con questo contributo intendiamo portare l'attenzione sul fatto che nel quadro del disordine informativo la macroarea della disinformazione rischia di fagocitare le informazioni attendibili, rappresentando una minaccia della democrazia, capace di inquinare le opinioni pubbliche e sfidare i sistemi democratici (Giusti, Piras 2023). La diffusione di informazioni distorte può avere infatti un ruolo tanto rilevante nella destabilizzazione della sfera pubblica (Habermas 2023) e nella manipolazione e polarizzazione delle visioni (Bail *et alii* 2018) e del dibattito interno da alimentare sentimenti di disaffezione verso le istituzioni, configurandosi come un *patogeno* del corpo politico.

¹ A conferma, si può considerare quanto documenta l'OECD (2021), ovvero che la disinformazione è riconosciuta quale problema pubblico dal 54% dei Paesi esaminati, i quali hanno implementato specifiche *policies* e azioni istituzionali.

Nutrendosi di stereotipi e immaginari che hanno già avuto successo e sono già radicati nelle culture collettive, essa può essere molto virale e minare in maniera sottile la credibilità delle istituzioni, operando un capillare deterioramento del meccanismo della fiducia collettiva.

L'insieme di questi stress tende a frantumare i *fondamenti solidali* della collettività e della convivenza. Condizioni di incertezza, come quelle dovute tanto a una carenza che a una eccedenza di informazioni, richiedono un difficile processo di riduzione della complessità, che investe sia le risposte che le aspettative fiduciarie riposte in attori istituzioni e collettivi (fiducia sistemica o istituzionale) o in attori individuali (fiducia interpersonale) (Mutti 1998; Luhmann 1989). Se la fiducia non si rinsalda, la società tende a disintegrarsi (Simmel 1984). La carenza di fiducia compromette, infatti, le importanti funzioni di regolazione e integrazione sociale all'interno dei sistemi democratici, e la politica deve chiedere ingenti sforzi, emotivi e cognitivi, alla popolazione di fronte a misure che, non trovando elementi di fiducia a sostegno, diventano inevitabilmente costringenti, come si è verificato durante la pandemia.

Pertanto, se il disordine informativo rappresenta un rischio effettivo per la sicurezza e la coesione dei sistemi democratici contemporanei, generando disorientamento e crisi della fiducia, attraverso il riscontro di alcuni attori istituzionali che a vario titolo hanno responsabilità nella gestione e nella produzione di informazioni cruciali per la sicurezza dei cittadini, ci chiediamo come tale rischio venga percepito e interpretato, soprattutto in relazione al campo della disinformazione strategica, o strettamente intesa; al problema dell'erosione della fiducia; alle tecnologie comunicative e alle strategie che vi incidono; ai limiti di governabilità del fenomeno e alle risposte di contrasto ritenute efficaci ma anche legittime nei sistemi democratici.

Punto di partenza del nostro discorso è che gli attori che contribuiscono al disordine informativo possono essere soggetti non statali, sia leciti (partiti e movimenti politici, aziende e società finanziarie, gruppi di interesse e di pressione, organizzazioni religiose, organizzazioni non-governative) che illeciti (gruppi terroristici ed eversivi, organizzazioni criminali, "poteri occulti", sette religiose estremiste) (Germani 2017). Ma possono essere anche soggetti statali, ovvero gli Stati con i loro apparati. Da parte di tutti questi diversi attori, l'informazione distorta può essere rivolta sia all'interno che all'esterno di ogni Stato e, quindi, utilizzata per influenzare l'opinione pubblica al fine di rafforzarne o indebolirne l'integrità, producendo effetti decisivi anche sui conflitti. Dall'esterno, l'obiettivo è quello di destabilizzare contesti nazionali e/o istituzionali "non amici", minandone la coesione interna, delegittimandone le istituzioni politiche, alimentando le divisioni interne, estremizzando le contrapposizioni esistenti, per influenzare i processi decisionali. Dall'interno, la diffusione di informazione distorta può concorrere a minare la reputazione delle istituzioni nazionali e può offrire il fianco alla propaganda esterna. Il problema della disinformazione, pertanto, si pone come effetto combinato di una comunicazione in cui convergono sia le strategie degli Stati "nemici" sia la comunicazione che si produce nei singoli Stati e che senza controlli si insinua al loro interno, nei gangli della democrazia (Chadwich 2013).

Obiettivo del lavoro è contribuire al dibattito, considerando, in particolare, l'informazione strategicamente distorta da attori statali esteri verso l'Italia e andando a comprendere come il disordine informativo venga interpretato dai principali soggetti istituzionali, se e in che termini venga ritenuto un fattore di incertezza politico-istituzionale e inquadrato come minaccia, quale sia il nesso con l'erosione della fiducia e le esigenze di governo del fenomeno nelle democrazie contemporanee.

In relazione a tale obiettivo, questo contributo si avvale delle risultanze empiriche emerse da una ricerca² realizzata nel 2023 sulle rappresentazioni degli attori istituzionali italiani coinvolti nell'osservazione e nell'azione di contrasto alla disinformazione. Tale delimitazione della ricerca consente di distinguere il livello istituzionale dell'individuazione del rischio disinformativo sia dalle dinamiche di produzione dei contenuti sia dai loro effetti sull'opinione pubblica. Inoltre, essa permette di spostare l'attenzione dal fenomeno in sé alla sua identificazione come problema per la sicurezza e per la coesione democratica.

² La ricerca *"La guerra ibrida e l'opinione pubblica come Centro di Gravità dei conflitti moderni. L'impiego dell'ambiente informativo e dei social network per influenzare le opinioni pubbliche proprie e avversarie, rafforzando o indebolendo la coesione interna ed esterna e generando effetti decisivi sui conflitti e le crisi"* (Gab 1/2023) è stata affidata ad Antonella Coco dall'Ufficio di Gabinetto del Ministro della Difesa italiano nel 2023.

Insieme a ciò, partendo da una riflessione sul sistema mediale, l'articolo indaga anche il potenziale ruolo di rafforzamento o di contrasto delle nuove tecnologie della comunicazione, con attenzione all'Intelligenza Artificiale. La riflessione sull'IA, in questo quadro, sebbene non costituisca un ulteriore livello di analisi empirica, è volta a problematizzare le trasformazioni dell'ambiente informativo, al fine di capire, nel governo della diffusione di informazioni distorte, quali possano essere le strategie e gli strumenti tecnologici da impiegare, fino a comprendere se l'IA può avere un ruolo nel coadiuvare azioni statali di contrasto e nell'arginare le derive in atto. Questa domanda appare particolarmente rilevante nei contesti democratici laddove gli Stati sono chiamati a garantire sicurezza e libertà democratiche, quindi una informazione plurale e autonoma.

IL DISORDINE INFORMATIVO COME MINACCIA PER LE DEMOCRAZIE OCCIDENTALI

La disinformazione è parte della comunicazione, che costituisce un aspetto di cruciale importanza per il potere politico, in tutte le sue fasi e a tutti i suoi livelli. Panebianco (1982: 81) sottolinea come «chi ha la capacità di distribuire, manipolare, ritardare o sopprimere le informazioni controlla una fondamentale area di incertezza, ha in mano una risorsa decisiva nei rapporti di potere». Tale risorsa riguarda i mezzi di comunicazione che possono essere direttamente soggetti al potere politico (come le radio e le tv di Stato, controllate dalle istituzioni politiche che sono veri e propri datori di lavoro dei giornalisti) oppure possono costituire un sistema autonomo e via via più influente, con il quale la politica interagisce.

Rispetto a queste dinamiche, il potere politico utilizza e spesso combina due strategie principali: crea canali di informazione propri (ufficialmente o ufficiosamente); cerca di blandire o esercitare pressioni e controlli sugli attori mediali che non gestisce direttamente (Costabile 2002).

Un'altra distinzione, nelle teorie del potere, riguarda il modo in cui informazione e mezzi di comunicazione sono gestiti dai regimi democratici e da quelli non democratici. I primi favorirebbero la circolarità delle forme di comunicazione e la presenza di una pluralità di soggetti che emettono e ricevono messaggi, includendo un'ampia gamma di opportunità comunicative che possono partire anche dal basso, comprese forme di persuasione e influenza. I secondi, monopolizzando la comunicazione e controllandone l'emittenza, la utilizzerebbero come specifica forma di potere e manipolazione, dato il suo carattere di unilateralità e la sua direzione univoca da chi la trasmette a chi la riceve.

Nel tempo, in entrambi i tipi di regime, i cambiamenti del sistema mediale hanno ampliato le possibilità di distorcere le informazioni, introducendo nuovi rischi di manipolazione delle opinioni, delle percezioni e degli atteggiamenti. Con lo sviluppo tecnologico e digitale, con la rete e l'avvento dei social media, la diffusione incontrollata delle notizie è, tuttavia, aumentata soprattutto nei regimi democratici (tanto da far parlare di "caos" e "disordine informativo"), caratterizzati da una crescente pluralità di attori dell'informazione (Quattrocioni, Vicini 2023), tra i quali *citizen journalist*, ma anche e soprattutto attori e gruppi di interesse che costruiscono "*disinfo-as-a-service*", come le agenzie di marketing politico, le pseudo-testate giornalistiche, o le vecchie testate giornalistiche "zombificate" e riadattate e nuove logiche, influencer ideologici o interessati a vendere prodotti, gruppi di *austroturfing* o finti gruppi *grassroots*.

Ci troviamo, dunque, di fronte a una serie di spinte che tendono a radicalizzare il disordine informativo. La rete ha eroso progressivamente lo spazio dei media tradizionali, cambiando le dinamiche relative al *newsmaking*. Mentre i media tradizionali si basano su una trasmissione unidirezionale, rivolta ad un pubblico indifferenziato, e svolgono la loro funzione sulla base della gestione istituzionale delle fonti, i social media promuovono una comunicazione sia bidirezionale e orizzontale che unidirezionale e diffusiva. Essa si articola in due momenti: quello di una produzione e condivisione di contenuti disintermediati tra individui, o da parte dei singoli verso la rete (*bottom up*), e quella di una produzione e gestione di contenuti inviati ad hoc dalle big companies (*top down*). Sulle piattaforme, tali contenuti da un lato sono inseriti e diffusi dai *follower* o altri personaggi che non sono veri e propri professionisti dell'informazione, ma esperti nel trasformare le notizie in storie accattivanti e affabulatorie che si prestano a trasmissioni virali potenti e istantanee (Giungato, Taddei, Affuso 2024). Dall'altro lato sono aggregati attraverso gli algoritmi in bolle chiuse all'esterno, per cui è difficile intervenire con il *debunking*, e che non

comunicano tra loro, per cui si indebolisce il dibattito pubblico. L'azione delle piattaforme si basa su un preciso *business model* in virtù del quale il sistema è orientato a massimizzare la quantità di tempo che gli utenti trascorrono on line e a potenziare la circolazione non dei temi di migliore qualità, ma di quelli che riscuotono più successo in termini di visualizzazioni e di *like*. All'aumento dell'attenzione e degli scambi da parte del pubblico si accompagna un aumento di valore economico e di appetibilità da parte della sfera pubblicitaria. Per questo sono premiate le polarizzazioni delle opinioni, il dibattito e i messaggi divisivi che generano audience e aumentano le connessioni (Quattrocioni, Vicini 2023).

In questi processi, vi sono diversi fattori e mezzi che spiegano come le persone trattengono e diffondono le notizie. Tra i fattori vi sono le credenze, le ideologie, le culture, gli immaginari maturati e radicati nei contesti di appartenenza, capaci di condizionare visioni collettive; tra i vari mezzi che incidono sul disordine informativo, significativi e fortemente distorsivi sono i c.d. "media sintetici" e *deepfake*, ovvero contenuti digitali (video, audio, immagini o testo) manipolati attraverso tecniche/tecnologie di *machine/deep learning* che agiscono attraverso i *confirmation bias*, l'omofilia e le *echo chamber* (Del Vicario *et alii* 2016) con le relative polarizzazioni. Oltre ai *deepfake*, video e audio, ampiamente implicati sono i social bot, le reti neurali per la generazione di testi artificiali, nonché le tecniche di profilazione rese possibili da un processo sempre più pervasivo di *datificazione* della realtà.

Le trasformazioni descritte danno il quadro di un ambiente comunicativo caratterizzato da elevata ampiezza nella circolazione di notizie e pseudo-conoscenze ma da riduzione dei filtri tradizionali, da una semplificazione delle questioni, e al contempo da una estremizzazione delle posizioni, ed entro cui il disordine informativo si diffonde attraverso una molteplicità di pratiche e strategie comunicative (Bentivegna, Boccia Artieri 2021)³, generando instabilità nella sfera pubblica e negli ambienti digitali e *social*.

LE ISTITUZIONI ITALIANE DI FRONTE ALLE MINACCE DISINFORMATIVE: CARATTERISTICHE E RISCHI PERCEPITI

A partire dal quadro concettuale tratteggiato, la ricerca, di tipo qualitativo, si è basata sull'analisi congiunta della letteratura, di fonti documentali prodotte a livello nazionale e sovranazionale (Ministero della Difesa, NATO, Unione europea) e di informazioni tratte da rilevazione empirica. Tali informazioni sono state raccolte attraverso la realizzazione di interviste ad attori istituzionali. Il fine è stato mettere in luce le loro percezioni e rappresentazioni del rischio associato alla disinformazione, nel contesto italiano, per comprendere sia gli aspetti in base ai quali viene interpretata come problema pubblico e come minaccia sia le strategie di risposta.

Nel complesso, sono state realizzate 12 interviste qualitative-semi-strutturate. Di queste, nel 2023 ne sono state somministrare 3, esplorative, a studiosi afferenti al mondo accademico e della ricerca; 6 agli attori istituzionali impegnati a vario titolo nell'osservazione e/o nel contrasto alla disinformazione; 3, nella fase finale di elaborazione delle informazioni, nel 2025, rivolte a giornalisti, divulgatori e analisti coinvolti nel fenomeno osservato. Gli intervistati sono stati selezionati secondo un criterio intenzionale, sulla base della loro posizione istituzionale e della rilevanza rispetto al fenomeno oggetto di studio e provengono dalle sfere della politica, della difesa, della comunicazione e della ricerca. I temi affrontati hanno riguardato i caratteri della minaccia disinformativa, gli attori interessati, l'evoluzione nel tempo, le vulnerabilità del nostro Paese e le strategie di governo del fenomeno. Il materiale empirico è stato analizzato mediante un approccio qualitativo di tipo tematico, orientato a corroborare la ricerca attraverso le loro interpretazioni ed esperienze.

³ Tra le tecniche e le strategie di disinformazione, oltre alla creazione di notizie non attendibili e *fake news*, ci sono il complottismo, la costruzione di hate speech, lo storytelling, l'affabulazione.

Tabella 1 – Profilo anonimizzato dei testimoni privilegiati intervistati (TP01–TP10).

TP01	Docente universitario 1
TP02	Docente universitario 2
TP03	Consigliere diplomatico Ufficio di Gabinetto Ministero Della Difesa
TP04	Dirigente Direzione Generale per Affari Politici e di Sicurezza - Ministero degli Affari Esteri e della Cooperazione internazionale 1
TP05	Dirigente Direzione Generale per Affari Politici e di Sicurezza - Ministero degli Affari Esteri e della Cooperazione internazionale 2
TP06	Giornalista - <i>Italian Digital Media Observatory</i>
TP07	C.do Gen. CC - II Reparto - Ros
TP08	Direzione Agenzia per la Cybersicurezza nazionale
TP09	Giornalista Rai
TP10	Analista ed esperto di comunicazione in contesti di crisi
TP11	Giornalista scientifico e divulgatore
TP12	Direttore Istituto di ricerca

L'analisi integrata della letteratura, delle fonti documentali e delle interviste a testimoni qualificati, che restituiscono il quadro entro cui gli attori istituzionali definiscono e interpretano il fenomeno, fa emergere che, nel contesto indagato, la disinformazione è considerata un “problema pubblico” (Lovari, Ducci 2022).

Come evidenziato da Gusfield (1984), un problema sociale acquisisce rilevanza pubblica quando diversi attori, le istituzioni e le autorità governative ne riconoscono l'esistenza, lo portano all'interno dell'agenda politica e iniziano a realizzare politiche di intervento, ideando possibili soluzioni.

Sebbene sia stata inizialmente oggetto di attenzione da parte di pochi esperti, la disinformazione, come testimoniano gli attori intervistati, è oggi parte dell'agenda dei governi a livello nazionale e internazionale. Già in seguito al referendum del 2016 nel Regno Unito sulla Brexit (Cassidy 2023), con lo scandalo *Facebook-Cambridge Analytica*, e in seguito all'utilizzo da parte della Russia di informazioni false come strumento di politica estera e parte integrante della dottrina militare di Putin (Comai 2023; Di Pasquale e Germani 2023; Zafesova 2023), la Commissione Europea ha manifestato preoccupazioni in merito all'impiego di “campagne di disinformazione di massa online” volte a generare sfiducia e tensioni nella società (European Commission 2022). Con la pandemia da Covid-19 il fenomeno si è amplificato ed è divenuto particolarmente visibile (Cinelli *et alii* 2020; Hazelton 2023). Oggi, soprattutto a partire dalle vicende legate alla guerra russo-ucraina, le istituzioni internazionali considerano la diffusione di informazioni false una minaccia tipica dei cosiddetti conflitti ibridi (*hybrid warfare*)⁴.

⁴ Il concetto di “conflitto ibrido” ha acquisito rilevanza in ambito Nato e nell'Unione Europea a partire dal 2014, in relazione ad eventi quali l'annessione della Crimea e la destabilizzazione dell'Ucraina orientale da parte della Russia, l'affermarsi di Daesh. La Nato definisce le minacce ibride, in senso ampio, come l'impiego di mezzi non convenzionali – tra cui l'utilizzo della propaganda, di risorse digitali, economiche, sociali o attacchi cyber – al fine di destabilizzare le società dei paesi avversari (Cusumano, Corbe 2018; Nato 2014).

Con riguardo all'Italia, la Relazione sulla "Politica dell'informazione e della sicurezza 2022" (Sistema d'informazione per la sicurezza della Repubblica) denuncia la conduzione di campagne aggressive rivolte ai Paesi occidentali, e anche all'Italia, «tentando di manipolare a proprio vantaggio gli scambi commerciali, le forniture di beni strategici e l'ambiente informativo dei Paesi bersaglio» (Ibidem). Inoltre, individua, tra le principali caratteristiche delle minacce ibride legate alla disinformazione, la difficoltà di attribuzione della comunicazione. In particolare, le operazioni disinformative sono descritte come difficili da identificare e da ricondurre a specifici attori, in quanto condotte in modo coordinato su più domini – diplomatico, militare, economico e mediatico – e spesso mantenute al di sotto della soglia del conflitto (Ibidem). Tale dinamica si traduce nella diffusione di contenuti attraverso canali "multi-tematici" che sviluppano narrazioni crossover, riportando notizie false e tendenziose su molti argomenti (Luiss 2023).

La Relazione indica Russia, Cina, Iran e Corea del Nord tra i principali attori statuali impegnati in strategie di influenza e destabilizzazione delle opinioni pubbliche. In particolare, evidenzia come Mosca e Pechino conducano campagne multivettoriali, sfruttando l'apertura delle società occidentali, nei mercati e negli ambienti mediatici, e collegandosi a reti più ampie di Paesi – tra cui Venezuela, Cuba e Siria.

Se ci si sofferma sulla Russia, si coglie un caso paradigmatico di come i media, nei regimi autoritari, costituiscano un pilastro delle strutture di potere statale. Il controllo dell'informazione è un elemento centrale del modello neo-autoritario putiniano, capace di limitare l'accesso a fonti plurali e produrre tra i cittadini russi un gap di consapevolezza rispetto a quanto avviene all'interno e all'esterno dello Stato (Antinori 2023). Proprio all'esterno, la strategia russa si sviluppa attraverso operazioni di *Foreign Information Manipulation and Interference* (FIMI), che combinano la distorsione dell'informazione e la propaganda e mirano a sfruttare le linee di frattura presenti nelle società occidentali per destabilizzare il tessuto sociale e rendere fragili le democrazie (ivi: 132). A partire dal 2014, con l'annessione della Crimea e il conflitto nel Donbass, la disinformazione è stata impiegata in modo sempre più sistematico come elemento delle strategie di conflitto, affiancando alle operazioni militari un'intensa attività di manipolazione informativa. Negli anni successivi, tali pratiche si sono estese a diversi contesti, incluse interferenze nei processi politici occidentali (come nelle elezioni presidenziali americane del 2016 e nel referendum della Brexit) e nella gestione comunicativa relativa a crisi globali (come la pandemia, in un progressivo aumento di intensità e sofisticazione). L'invasione del 2022 dell'Ucraina da parte della Russia ha segnato l'inizio di una guerra dell'informazione ancora più intensa e visibile agli occhi del mondo, con l'attivazione di tutti i canali propagandistici da parte del Cremlino per giustificare l'aggressione e mantenere il sostegno interno. L'obiettivo complessivo è apparso quello di destabilizzare le società democratiche, rendendo più fragile la capacità di costruire orientamenti condivisi.

Negli ultimi anni, anche la Cina ha intensificato le campagne di influenza mediatica all'estero. Se prima del 2020 mirava alla promozione di un'immagine positiva del Paese, durante la pandemia di Covid-19, in convergenza con la Russia, si è impegnata anche a diffondere teorie cospirative e disinformazione su scala globale, seminando confusione e dubbio nei cittadini occidentali riguardo alle indicazioni delle proprie autorità sanitarie. L'obiettivo principale appare quello di proteggere e rafforzare la reputazione del regime, anche attraverso la negazione o la minimizzazione di questioni sensibili, come Hong Kong, Taiwan o le violazioni dei diritti umani.

Su questa base, dall'analisi delle interviste emergono alcuni temi ricorrenti che rendono conto di come gli attori istituzionali definiscono e interpretano la disinformazione non solo come problema informativo ma come una questione che riguarda la sicurezza e la coesione democratica e che richiede nuove forme di governance.

La rete, per loro, ha reso più agevole la diffusione di contenuti falsi, favorendone l'utilizzo anche nell'ambito dei conflitti tra Stati. Tuttavia,

rispetto al giornalismo mainstream, non ha costituito soltanto uno spazio di *citizen journalism* e di diffusione dei valori democratici ma è diventata anche l'ambito in cui condurre la guerra ibrida della disinformazione (TP06)

Gli intervistati rilevano, inoltre, l'esposizione dell'Italia a minacce di disinformazione provenienti da attori esterni, pur in assenza di eventi specifici rivolti al nostro Paese tali da configurare un pericolo

immediato per la sicurezza nazionale. Essi individuano due momenti particolarmente rilevanti in termini di condizionamento delle opinioni e sottolineano di essere

rimasti esterrefatti nel vedere con quanta rapidità le fonti, i siti e le personalità che erano impegnati in altre informazioni sul COVID-19 e sui vaccini, il 22 Febbraio del 2022 sono passati a fare disinformazione sull'Ucraina. (...) quello che è accaduto è che personalità, fonti, siti, reti che non si erano mai occupati di conflitti nell'Europa orientale passavano a inneggiare Putin e deprecare contro Zelenski. (TP 06)

I nostri interlocutori evidenziano anche il ruolo delle tecnologie digitali nell'ampliare il potere della disinformazione, rendendo più rapida e pervasiva la diffusione delle notizie e accrescendo la capacità di influenzare sia l'opinione pubblica sia i decisori politici.

Quanto all'IA, sottolineano che il suo sviluppo consente una crescente sofisticazione dei contenuti informativi, rendendo sempre più difficile distinguerli dalle informazioni attendibili.

La disinformazione cambierà ancora, sembrerà sempre più informazione reale perché gli strumenti tecnologici stanno sviluppando capacità che rendono determinate informazioni sempre più veritiere, per cui aumenta la capacità di condizionare l'opinione pubblica. (TP 07)

Gli osservatori sottolineano la potenza della disinformazione come strumento utilizzato nei conflitti contemporanei tra Stati.

L'idea per cui pur non essendo un combattuta in armi è una guerra ha aperto e preceduto il conflitto in Georgia nel 2006, ha aperto e ha accompagnato il conflitto in Siria e in Ucraina nel 2014, nel 2022 il conflitto in Ucraina, sta accompagnando il conflitto in Medio Oriente tra Hamas e Israele. (TP 06)

Quella della disinformazione appare dunque come una minaccia permanente, continuativa nel tempo, seppure con livelli di intensità diversa, senza una vera e propria distinzione fra periodi di pace o di guerra, inscritta stabilmente nell'ambiente informativo contemporaneo. Essa, inoltre, si caratterizza per ambiguità e incertezza rispetto alla sua stessa esistenza, per cui chi riceve tale minaccia non sempre la percepisce in maniera chiara e non ha la certezza di essere colpito.

Un'altra peculiarità indicata riguarda il fatto che non è sempre possibile conoscere la provenienza delle informazioni, rendendo così difficile l'attribuzione della comunicazione agli attori responsabili. La possibilità che i messaggi siano veicolati da attori terzi rispetto alle parti in conflitto aumenta queste difficoltà.

La caratteristica di extraterritorialità rende difficile capire da dove viene, da dove si origina la comunicazione quindi fare un'attività che possa essere di carattere di contrasto effettivo a qualche informazione è difficilissima. (TP07)

La rappresentazione della disinformazione come minaccia sistemica e permanente si riflette nella seconda dimensione emersa dall'analisi: l'erosione della fiducia. Infatti, le caratteristiche che qualificano la disinformazione – la sua continuità nel tempo, la difficoltà di attribuzione e la crescente sofisticazione dei contenuti – contribuiscono a produrre una condizione di disorientamento nello spazio informativo e di incertezza nella distinzione tra fonti attendibili e non attendibili, che, a sua volta, incide sui meccanismi di costruzione della fiducia su più livelli (Intervista TP 11). Ne deriva una sfiducia che coinvolge non soltanto la dimensione strettamente informativa ma anche quella politico-istituzionale e quella relazionale. La manipolazione delle informazioni, allo scopo di colpire la coesione sociale, è costruita in modo tale da alimentare le divisioni interne, estremizzando le contrapposizioni, fino a generare sentimenti di avversità, odio e paura. In alcuni casi, tale dinamica può assumere forme più radicali che investono i meccanismi di costruzione del legame sociale, dando luogo a un atteggiamento che gli interlocutori definiscono

“nichilista”, caratterizzato da una generalizzata perdita di fiducia e dalla difficoltà a riconoscere qualsiasi fonte come affidabile e credibile.

Non credi più a nessuno, entri in una postura intellettuale in cui diffidi di tutto e potrebbe sembrare una posizione salutare, diffidare di tutto, ma quando la diffidenza diventa nichilismo cioè non è che diffido di tutto per credere a pochi ma diffido di tutto per non credere a nessuno, allora (...) si va a quella deriva di polarizzazione populista, quella deriva di imposizione astiosa, rabbiosa per cui il dialogo diventa impossibile, le società si dividono. (TP06)

La disinformazione tende ad accentuare la segmentazione dello spazio pubblico e rafforzare visioni contrapposte difficilmente conciliabili.

Si va a quella deriva di polarizzazione populista, quella deriva di imposizione astiosa, rabbiosa per cui il dialogo diventa impossibile, le società si dividono [...] quando attorno alla politica il consenso e il negoziato, la trattativa, il compromesso è un tradimento. (TP 06)

Gli intervistati, inoltre, riconoscono la sfiducia non soltanto come effetto della disinformazione, ma anche come uno dei principali obiettivi degli attori avversari. In altri termini, la distorsione dell'informazione è rappresentata come una strategia orientata alla produzione intenzionale di sfiducia, quale meccanismo di indebolimento della coesione democratica. Le campagne disinformative, infatti, non mirano esclusivamente a persuadere rispetto a specifici temi, ma anche a generare una condizione più ampia di delegittimazione: «l'obiettivo non è creare un dubbio su questa tesi o questa notizia o questa personalità, l'obiettivo è segnare sfiducia» (TP 06).

Quanto delineato fin qui comporta, secondo i testimoni qualificati, la necessità e al tempo stesso le difficoltà nel governare il fenomeno della disinformazione. Un sistema democratico è chiamato a tutelare la libertà di pensiero e di espressione, evitando derive di tipo securitario, ma al tempo stesso a governare il rischio connesso alla circolazione dei contenuti falsi, in un equilibrio che implica «la tutela della libertà accompagnata dalla conoscenza» (TP 03).

In Europa, alcuni Paesi hanno istituito organismi dedicati al contrasto alla disinformazione. Nel caso italiano, gli intervistati non reputano necessaria né sufficiente l'istituzione di un organismo istituzionale specifico. Tale soluzione richiederebbe un ingente impiego di risorse per la gestione dei flussi informativi senza garantire una maggiore capacità di risposta né consentire una chiara misurazione dei risultati (TP 08).

Emerge una preferenza per approcci orientati a forme di governance diffusa e al coinvolgimento di una pluralità di attori appartenenti a diverse sfere della società – politico-istituzionale, economica, mediatica e culturale – includendo, quindi, tra gli altri, organi governativi, media, sistema educativo e organizzazioni della società civile. Prevale l'idea secondo cui il contrasto alla disinformazione richiede un approccio complessivo, noto come *whole-of-society approach*⁵, volto alla costruzione di società resilienti, attraverso il rafforzamento delle capacità di analisi critica (TP 04 e TP 05). Secondo questa prospettiva uno dei goal delle politiche è proprio quello della costruzione della fiducia istituzionale: «Allora dobbiamo ricostruire fiducia, ma questo è un compito molto complesso» (TP 06)

In questo modello di governance diffusa, la sfera mediatica assume un ruolo centrale. Anche il rafforzamento dell'informazione plurale è considerato un aspetto fondamentale. I nostri interlocutori sottolineano il ruolo fondamentale dei media tradizionali in termini di responsabilizzazione e professionalizzazione dell'attività giornalistica, riaffermando la rilevanza della re-intermediazione rispetto agli spazi di *citizen journalism*. Al tempo stesso, emerge una diffusa consapevolezza dei limiti nel contrasto alla disinformazione delle sole operazioni di *fact-checking*, *debunking* e *prebunking*. L'aumento dei contenuti e la capacità della disinformazione di produrre confusione rendono tali strumenti insufficienti, sia per la mole di informazioni da verificare, sia per le difficoltà legate alla selezione e all'interpretazione dei

⁵ L'espressione “*whole-of-society approach*” è utilizzata in ambito di policy, in particolare nei documenti NATO e dell'Unione europea, per indicare un modello di governance delle minacce basato sul coinvolgimento coordinato di una pluralità di attori pubblici, privati e della società civile, e finalizzato al rafforzamento della resilienza dei sistemi democratici (NATO 2021; European Commission 2020; OECD 2020).

contenuti (Intervista TP03). Inoltre, quando tali attività sono svolte a livello politico-istituzionale, emergono rilevanti questioni in termini di legittimità e controllo democratico: chi definisce i criteri di intervento? Come è possibile garantire un controllo democratico su queste attività? Chi controlla il “controllatore”? Si tratta di interrogativi che, nei sistemi democratici, risultano centrali nell'affrontare il problema.

Accanto alla sfera mediatica, quella culturale – e in particolare il sistema dell'istruzione – riveste un ruolo rilevante nel contrasto alla disinformazione. Gli intervistati sottolineano come tale contrasto non possa limitarsi alla *media literacy* ma richieda un più ampio innalzamento del livello di istruzione e delle capacità critiche. Incrementare il livello di conoscenze e di consapevolezza della popolazione e accrescere la cultura digitale rappresentano obiettivi strategici nel contrasto alla disinformazione (Intervista TP 08). In questa prospettiva, gli investimenti nell'istruzione pubblica sono indicati come strumenti fondamentali.

Emerge poi la necessità di una governance sovranazionale del fenomeno della disinformazione, che, per la sua natura transnazionale, prescinde dai confini dei singoli Stati. I testimoni qualificati sottolineano, a tal proposito, il ruolo delle piattaforme online nelle politiche di moderazione e limitazione dei contenuti e degli utenti. In questa direzione, vengono valutate positivamente le iniziative dell'Unione europea volte a promuovere forme di co-regolamentazione dello spazio digitale, basate sul coinvolgimento diretto e sulla responsabilizzazione delle piattaforme.

DISINFORMAZIONE ARTIFICIALE E (S)FIDUCIA ISTITUZIONALE

Con l'ingresso nella fruizione di massa delle piattaforme digitali e, in quest'ultima fase, dell'IA, la disinformazione, come si coglie anche dalle testimonianze degli attori intervistati, riesce a essere ancora più mirata e difficilmente riconoscibile, grazie alla crescente capacità delle tecnologie contemporanee di imitare il linguaggio e la realtà.

A renderlo palese è uno dei primi contenuti generati allo scopo di sensibilizzare il pubblico sulla potenza dell'IA nel manipolare volti, corpi o voci, e simulare eventi mai accaduti tramite e il video sottolineare la facilità con cui un messaggio falso può apparire autentico e virale. È il *deepfake* politico creato da BuzzFeed News, nell'aprile 2018, in collaborazione con il regista Jordan Peele. Due sistemi (un generatore e un discriminatore) si sfidano in un processo ripetitivo, come in una sorta di “duello tra falsario e detective”. Esso mostra l'ex Presidente Barack Obama in un apparentemente consueto annuncio istituzionale. Inizialmente lo spettatore vede solo il Presidente Obama, che pronuncia frasi mai dette, tra cui un'offesa al presidente Donald Trump. Ma poi scopre che la voce e il labiale, perfettamente sincronizzati, sono quelli dello stesso Peele.

Il video è stato pensato per sollevare interrogativi cruciali sull'affidabilità dei documenti digitali nel campo politico, sul ruolo di una tecnologia ampiamente accessibile, economica e in rapida diffusione, sui rischi concreti per l'integrità del discorso pubblico e per la sicurezza democratica⁶. Si può dire che a partire dal momento della sua diffusione, gli effetti della tecnologia sugli avvenimenti e le campagne disinformative del 2016, insieme al ruolo giocato dalle grandi piattaforme digitali e dai loro imprenditori (ad esempio Facebook e Zuckerberg prima, X e Musk dopo), diventano chiari anche per il grande pubblico. Si capisce soprattutto come il fenomeno della disinformazione sia più articolato, rispetto al problema delle sole *fake news*, e come tale articolazione abbia impatti importanti rispetto agli orientamenti collettivi, compresi quelli di voto.

Dopo quasi dieci anni, una gran mole di rapporti e analisi sulla disinformazione evidenziano che seppure essa «non rappresenti un'assoluta novità strategica, le innovazioni tecnologiche e le modifiche nell'ecosistema delle informazione degli ultimi anni hanno aggiunto ulteriori livelli di complessità, portando allo sviluppo di nuove e più sofisticate tattiche e strategie, che rendono queste campagne più rapide ed efficaci e permettono a diversi gruppi di attori, tanto pubblici quanto privati, tanto domestici quanto stranieri, di intervenire in maniera ingannevole nel dibattito democratico» (Luiss 2023: 6). La

⁶ Il video è attualmente disponibile su <https://www.rainews.it/video/2025/09/lattore-jordan-peelee-parla-ma-noi-vediamo-obama-il-video-deepfake-del-2018-9ef0335e-33e1-40b8-8e9c-473303e23ea7.html>, consultato il 27 settembre 2025.

potenza di questi contenuti nell'influenzare l'opinione pubblica e nel condizionare processi decisionali può essere quindi tanto alta da indebolirne le principali funzioni: da un lato quella relativa all'esercizio deliberativo, dall'altro quella relativa al controllo sulle dinamiche di potere (Bonfanti 2021).

Dunque, è in questo panorama tecnologico che si pone il problema della vulnerabilità delle democrazie occidentali contemporanee per le minacce riconducibili a campagne di disinformazione promosse da Stati avversari e finalizzate a creare instabilità e a indebolirne la coesione interna. Sono minacce legate a strategie e tecniche di manipolazione dell'opinione pubblica, a una tecnologia sofisticata che alcuni studi recenti considerano come una nuova sfida per i contesti democratici, che ne possono risultare particolarmente danneggiati (Luiss 2023).

Ad incidere è in grande misura anche la diffusione di discorsi che sfruttano e stimolano atteggiamenti discriminatori, incitano sentimenti di odio, violenza e possono sfociare nell'avversione e nel terrorismo. Se la propaganda ha sempre avuto l'obiettivo di mobilitare il sostegno rafforzando la propensione ai discorsi divisivi e l'appoggio al conflitto, costruendo un nemico, oggi è cambiata la velocità, la scala, la precisione, e la stessa possibilità di mistificare il processo attraverso cui i messaggi possono essere diffusi.

Favorite dalla tecnologia, tali speculazioni "informative" producono emozioni e opinioni virali, accrescono la confusione e l'incertezza. E questo può innescare fenomeni di sfiducia nelle istituzioni politiche e nelle agenzie giornalistiche, accompagnati all'emergere di sentimenti di paura e di ostilità verso e tra persone e/o gruppi considerati diversi (Giusti e Piras 2023), come anche l'emergere di soggettività e personalità capaci di assumere rilevanza sul piano politico attraverso la destabilizzazione delle opinioni pubbliche.

La fiducia può essere intesa come una risorsa relazionale (Mutti 1998) che consente agli individui di orientarsi in contesti caratterizzati da complessità, rendendo possibile l'adesione a decisioni e ordinamenti anche in condizioni di informazione incompleta (Luhmann 2002). Essa rinvia ad aspettative di affidabilità, prevedibilità e correttezza che si consolidano nel tempo e che sostengono il rapporto tra cittadini e istituzioni. Dal momento che la fiducia è essa stessa necessaria a fronteggiare l'incertezza, il destinatario di aspettative fiduciarie può essere costituito dall'organizzazione istituzionale e collettiva o da attori individuali (Mutti 1998).

La disinformazione agisce come fattore di erosione della fiducia istituzionale (Lazer *et alii* 2018) poiché interviene sui presupposti cognitivi, emotivi e simbolici che rendono possibile la formazione di aspettative affidabili. La crescita del disordine informativo e in particolare la circolazione di contenuti ambigui, falsi o manipolati compromettono infatti la riconoscibilità e attendibilità delle fonti, la credibilità delle voci istituzionali. Ne deriva una condizione in cui la fiducia diventa più difficile da tutelare e stabilizzare, mentre si rafforzano dinamiche di sospetto, distanziamento e delegittimazione nei confronti degli attori della democrazia. Questa degradazione si manifesta su alcuni livelli, distinti ma interconnessi, secondo una dinamica circolare di influenza reciproca: quello istituzionale e quello interpersonale (Mutti 1998). Tanto la fiducia nelle fonti dell'informazione che la fiducia verso la politica rientrano nella fiducia istituzionale, mentre a livello interpersonale ad essere coinvolta è la fiducia tra i cittadini e verso l'altro.

Sul piano dell'informazione, la creazione di notizie false, in vari formati, audio, video, testuale, ma sempre più indistinguibili da quelle attendibili (con testate zombie e citazioni che complicano la verifica, usate per sostenere verità politiche, economiche, sanitarie di grande portata), erode la fiducia pubblica nelle fonti e nell'autorevolezza dei professionisti e delle istituzioni che operano in questo campo, come i giornalisti e le varie testate. Nel complesso, la sfera pubblica e democratica ne viene depotenziata, ritrovandosi esposta e indifesa rispetto a interventi manipolativi sofisticati che ne compromettono la discorsività e decisionalità (Habermas 2023; Bonfanti 2021).

Per quanto riguarda la fiducia politica, ovvero la fiducia dei cittadini nel fatto che gli attori e le istituzioni politiche agiranno nel loro interesse, si palesa il rischio che la disinformazione porti con sé un calo progressivo della fiducia nelle istituzioni *rappresentative* (parlamenti, governi, partiti), a fronte di una tendenziale stabilità, della fiducia nelle istituzioni *esecutive* e amministrative (servizio civile, polizia, sistema giudiziario). Si tratta di una divergenza che può diventare terreno fertile per le campagne di disinformazione, deputate non necessariamente all'attacco dello Stato, ma al processo politico, incidendo in maniera faziosa su elezioni, dibattiti politici, funzionari eletti, dipinti come corrotti e insensibili ai bisogni dei cittadini. L'obiettivo è instillare la convinzione che "il sistema sia truccato", anche se le

funzioni quotidiane dello Stato sembrano funzionare. Questa strategia può allontanare la popolazione dalla politica, creando uno scenario in cui i cittadini possono fare affidamento sullo Stato per i servizi, ma non sentono alcuna lealtà verso il sistema democratico e legislativo. Si tratta di una condizione di precollasso, ottenuta con un investimento minimo da parte di chi intende paralizzare i nervi del potere democratico, parafrasando Deutsch (1962), lasciando il “corpo” amministrativo intatto ma senza direzione. Questo mina la fiducia nei processi democratici, attraverso retoriche note come quelle sull'integrità delle elezioni, e il tempo stesso delinea un campo di controllo per i media.

Accanto alla fiducia politica, la disinformazione può erodere un altro pilastro fondamentale: la fiducia verso l'altro. A essere intaccata è la convinzione nell'onestà, integrità e affidabilità degli altri, una “fede nelle persone” che non si conoscono personalmente. Dal punto di vista sociologico, è essenziale per controbilanciare queste spinte la cooperazione, la tolleranza, l'impegno civico e il funzionamento di una società complessa. La disinformazione distrugge la fiducia interpersonale alimentando la divisione, l'ostilità, la diffidenza. Attraverso la costante amplificazione di narrazioni “noi contro loro”, la rappresentazione degli avversari politici come minacce esistenziali e la promozione di teorie del complotto, essa incoraggia i cittadini a guardarsi l'un l'altro con sospetto e ostilità. Mostra i concittadini come nemici irrecuperabili impedendo che vengano invece vissuti come persone con opinioni diverse ma interlocutori legittimi, con i quali costruire intesa e forme di controllo del potere democratiche, non ostili, non violente.

UNO SGUARDO AL GIANCO BIFRONTE. L'IA TRA POTENZIAMENTO E CONTRASTO ALLA DISINFORMAZIONE

Abbiamo fin qui posto attenzione a come l'uso delle tecnologie della comunicazione e dei social media, insieme all'emergere di nuove narrative politiche disinformative, costituiscano, oggi, una sfida (per lo più minacciosa) per le democrazie contemporanee (Giusti, Piras 2023). In questo quadro, l'IA sembra configurarsi come un potente moltiplicatore del disordine informativo; il suo impatto si concretizza attraverso mediazioni legate alle modalità di implementazione, ai contesti socio-politici in cui è inserita e alle percezioni che ne influenzano gli effetti (Jungherr 2023).

Sebbene *rumors* e propaganda abbiano sempre accompagnato la comunicazione umana, specialmente nei conflitti, la capacità odierna di diffondere messaggi su vasta scala costituisce un elemento propulsore inedito. L'ascesa dei social media ha reso possibile la condivisione istantanea di dati tra milioni di persone, e l'IA ha integrato questa potenzialità, permettendo al pubblico di generare ed attingere ad infiniti contenuti, connotati da una plausibilità inedita.

Nell'odierno ecosistema mediale, la disinformazione non è solo più veloce, ma anche più mirata e difficilmente riconoscibile, o controllabile. Tecnologie come *deepfake* (video e audio), *social bot* e reti neurali per la generazione di testi sintetici sono tra i maggiori strumenti. Essi consentono di creare account automatizzati che simulano il comportamento umano, interagiscono in tempo reale e contribuiscono all'amplificazione virale di contenuti di ogni genere. Il risultato è un ecosistema inquinato, in cui la percezione del consenso pubblico può essere costruita artificialmente e l'opinione pubblica influenzata con metodi subdoli ma raffinati, anche in virtù di logiche algoritmiche orientate al coinvolgimento degli utenti e radicate nel modello di business delle piattaforme (Bontridder e Pouillet 2021).

Non si può in questo quadro sottovalutare il punto a cui sono giunte le IA generative. Fondate su Large Language Models (LLM), ovvero su insiemi vastissimi di testi, perlopiù raccolti alla rinfusa dalla rete e che un software organizza in base alla ricorrenza statistica delle parole, esse generano contenuti originali e convincenti nella forma ma non necessariamente corrispondenti a dati reali. Ciò avviene in virtù dello stesso meccanismo generativo dell'IA, da cui originano le cosiddette “allucinazioni”, che non sono errori in senso classico ma output semanticamente plausibili, che rappresentano l'esito di una progettazione. Le IA infatti non sbagliano, eseguono. Se, ad esempio, una calcolatrice dicesse che il risultato di 2 per 2 è 5, indicherebbe un malfunzionamento della macchina. Nel caso dell'IA generativa, se inventa una data che non esiste non sta commettendo un errore: sta usando la sua capacità di colmare

un vuoto, di dare l'informazione richiesta. Interagendo con un LLM si interroga quindi un sistema che è stato addestrato per generare testi coerenti, che tuttavia non conosce la realtà, non la comprende né la spiega. Un LLM produce sequenze di parole, predice la parola successiva, costruisce frasi in base a un addestramento e alla probabilità che un insieme di vocaboli che si susseguono sia coerente in un sistema linguistico, ma non verifica. Ciò ha una conseguenza fondamentale: su un contenuto elaborato in tal modo non si può fare affidamento per compiere scelte e azioni ponderate.

Se questo è il contesto generale, in ambito politico l'uso dell'IA per condizionare l'opinione pubblica è ormai una pratica diffusa. Un esempio emblematico è rappresentato dai cosiddetti *cyber troops*, definibili come attori statali o di partito che conducono campagne di propaganda online mediante strumenti automatizzati. Queste operazioni includono la creazione di account falsi, la diffusione di messaggi d'odio e l'uso di tecniche di *micro-targeting* basate sulla raccolta, spesso illecita, di dati personali per personalizzare i messaggi persuasivi. Oltre al già menzionato caso di *Cambridge Analytica*, che ha svelato come i dati di milioni di utenti Facebook siano stati usati per influenzare il referendum sulla Brexit e le elezioni presidenziali americane del 2016, in diversi altri paesi, non solo occidentali, come Brasile, India e Nigeria, sono stati impiegati account automatizzati a sostegno di campagne elettorali recenti, o per screditare movimenti antagonisti al potere, come in Cina.

Tuttavia, se l'IA può amplificare la diffusione di contenuti non veri, potrebbe anche rappresentare una risorsa nel contrasto alla disinformazione (Kerystova 2018). La stessa infrastruttura tecnologica che alimenta i modelli generativi può essere orientata a costruire sistemi di rilevamento di contenuti manipolatori. Modelli linguistici avanzati, basati sul Natural Language Processing, o architetture ibride, che combinano per esempio BERT di Google e LSTM, o ancora GPT-4, di OpenAI, possono essere impiegati per classificare testi e integrare l'analisi semantica con quella strutturale delle notizie false, oppure analizzare, con le capacità multimodali, input visivi e sonori, per affrontare la complessità della disinformazione contemporanea su diversi canali comunicativi.

Tra le tecniche, *Zero-Shot Classification* consente a un modello di classificare contenuti mai visti durante l'addestramento. Questa capacità è cruciale, poiché le *fake news* si presentano in forme sempre nuove, rendendo obsoleti gli approcci tradizionali. Un modello *zero-shot* può, ad esempio, riconoscere una teoria complottista basandosi su una descrizione generale. Vi sono poi strumenti come DetectGPT capaci di rilevare se un testo è stato generato da un'IA. Infine, esistono piattaforme come *Tanbih News Aggregator* del *Qatar Computing Research Institute*, che, a differenza del fact-checking tradizionale, che interviene dopo la diffusione di una notizia, analizzano l'affidabilità e la polarizzazione delle fonti *prima* che gli articoli vengano letti. Valutando il profilo linguistico, l'attività social e la reputazione di una testata, il sistema fornisce un profilo dettagliato che aiuta l'utente a navigare consapevolmente tra le fonti, uscendo dalla propria "bolla" informativa e stimolando una lettura critica.

In base a questi strumenti, si tratta però di capire se e in che modo l'IA possa realmente aiutare a "valutare" l'attendibilità di una fonte. Da qualche anno, questa domanda è al cuore di vari studi (ivi). Più recentemente Walter Quattrocioni, insieme a Edoardo Loru, Jacopo Nudo, e colleghi (2025), in una ricerca mirata hanno chiesto a gruppi di valutatori umani ed a 6 dei principali modelli oggi in uso (Deepseek V3, Gemini 1.5 Flash, GPT-4o mini, Llama 3.1 405B, Llama 4 Maverick, e Mistral Large 2) di giudicare la credibilità di centinaia di siti di informazione, e classificarli come affidabili/non affidabili – motivando la scelta. I risultati mostrano come la questione centrale non riguardi l'accordo o il disaccordo su quali fonti risultino affidabili o meno, ma il modo in cui il "giudizio" viene effettivamente eseguito. E questo si evince sia quando i modelli e gli esperti concordano (per esempio su quali siti sono inaffidabili concordano anche oltre all'85%) che quando non concordano (per esempio gli LLM hanno classificato come non affidabile quasi l'80% delle fonti che gli umani considerano affidabili). In entrambi i casi, le motivazioni degli LLM si basano su criteri appresi dai dati di training che sono talmente distanti da quelli che orientano la valutazione umana da non poter essere considerati ragionevoli.

L'IA consente di rilevare l'uso di un certo tipo di parole, titoli *clickbait*, termini polarizzanti. Ma non potrebbe valutare gli obiettivi manipolatori, distinguere gli intenti, gli orientamenti, gli errori dalle strategie, le influenze dei contesti (Kerystova 2018). Agli LLM si possono delegare *framework* guidati da associazioni lessicali e statistiche piuttosto che da ragionamenti contestuali e da criteri normativi.

A fronte di tale processo, non appare più nemmeno sufficiente intervenire attraverso una governance algoritmica, capace di superare l'opacità di sistemi il cui controllo è in mano a poche piattaforme private che sfuggono a una reale responsabilità democratica. Anche qualora attraverso l'IA si riuscisse a controllare, e ottenere, unicamente l'immissione di dati attendibili, accertabili tramite fonti esperte, e di cui sia pubblico il controllo, e si giungesse a un vero e proprio costituzionalismo tecnologico che disciplini l'uso per tutelare diritti fondamentali e pluralismo informativo, ci si troverebbe di fronte a un sistema che quando produce una spiegazione, o una previsione, sta unicamente impiegando un'istruzione (ivi).

Se fino ad ora, dunque, il problema da gestire è stato quello del controllo dell'input dei dati in rete e della loro processazione (Affuso, Coco 2019), insieme a quello della diffusione algoritmica, oggi vi si aggiunge il fatto che l'IA genera uno spazio semantico che mima la facoltà umana di compiere valutazioni attraverso l'interpretazione del dato. All'IA non si chiedono informazioni, vere o false, ma valutazioni, sulla base delle quali prendere decisioni. E questo non può accadere.

La sfida affinché l'IA possa sostenere il contrasto al disordine informativo esula, quindi, sia dal problema di una supervisione della disinformazione come risultato, output, che dal problema di controllare le fonti nella fase dell'immissione, o input. Sia gli output che le fonti verrebbero analizzati nella ricorrenza statistica e qualunque risultato verrebbe restituito dal software come plausibile. Definiti «pappagalli stocastici» dalla linguista Emily Bender (2021) gli strumenti dell'IA eccellono nel risultare «intelligenti» e autorevoli. E lo saranno ancor più nella loro evoluzione prossima, quando non saranno più deputati a rispondere a un prompt per generare un testo, ma a comportarsi come «agenti» che compiono scelte su base modulare a partire da fonti e risposte selezionate, che svolgono compiti autonomi, come valutazione medico-legale, analisi di policy, servizi al consumatore.

Quello che diventa essenziale, di fronte alla possibilità di monitorare la disinformazione attraverso l'IA e di fronte ad output che possono condurre a risultati considerati utili, è valutare come le procedure interne operazionalizzano categorie normative, come l'affidabilità, e interagiscono con i *bias* umani. La vera questione non è se l'uomo sia «dentro» o «fuori» dal processo decisionale automatizzato, ma *come* vi partecipi, in quale momento della decisione, con quale grado di responsabilità e visione di società, insieme a chi.

Divenuta interlocutore credibile, l'IA rappresenta un ulteriore competitor rispetto a processi di costruzione del sapere basati sul vaglio pluralistico di soggetti istituzionali, competenti e capaci di sovrintendere a una informazione etica attraverso processi discorsivi collettivi, scientificamente verificati e democraticamente fondati. Con la sua natura tecnocratica, computazionale e numerica essa rischia di rendere ancora più profonda la crisi della sfera pubblica argomentativa e deliberativa, basata sul confronto tra individui in carne e ossa capaci di ragionare insieme, rendendo più debole la loro possibilità di monitorare i processi decisionali e sostenere scelte democraticamente e normativamente fondate (Habermas 2023). Ancorché mezzo tecnico utile, l'IA può sostenere uno sviluppo in senso democratico della sfera informazionale se l'innovazione che promuove viene bilanciata in base a modelli umano-centrici, in cui la macchina agisca come una sorta di copilota, capace di potenziare le capacità umane di salvaguardare la fiducia nel sapere e nella conoscenza.

CONCLUSIONI

Quanto proposto fin qui ci porta a una conclusione ancora del tutto parziale ma sostanziale. Il disordine informativo è interpretabile come una minaccia concreta ed esistenziale per le democrazie contemporanee, perché propone contenuti che si prestano a prendere di mira l'elemento fondante delle istituzioni: la fiducia.

Dall'indagine condotta, la prima dimensione tematica che emerge e ricorre nelle rappresentazioni degli attori istituzionali è che la disinformazione creata da soggetti statuali esteri verso il nostro Paese sia una minaccia permanente, non episodica né confinata a crisi specifiche, che si insinua in modo pervasivo nelle fratture preesistenti, sfruttando la crescente opacità dell'infosfera digitale, rendendo difficile individuare le fonti e attribuire responsabilità.

In questo quadro si colloca la seconda dimensione analitica emersa dall'indagine, relativa alla disinformazione come erosione della fiducia, che investe contemporaneamente la credibilità degli organi di informazione, la legittimazione delle istituzioni, la tenuta delle relazioni sociali, in un processo di reciproca alimentazione. L'indebolimento della fiducia costituisce al tempo stesso un esito della disinformazione ed un obiettivo perseguito dagli attori avversari nell'ambito dei conflitti con la finalità più ampia di colpire la coesione delle società democratiche. Il rischio è che si produca una condizione di sfiducia diffusa, istituzionale e interpersonale, che si nutre di diffidenza, astio, paura e mina le possibilità stesse del dialogo pubblico.

Le caratteristiche di pervasività, opacità e continua evoluzione del fenomeno osservato incidono sulla stessa difficoltà di governarlo. È la terza dimensione analitica messa in rilievo dalla ricerca, dalla quale emerge la necessità di una strategia di contrasto alla disinformazione. Quest'ultima non può essere affidata ad approcci securitari né a soli meccanismi di *fact-checking* e *debunking*, che appaiono in larga parte reattivi e insufficienti.

Può invece basarsi su un approccio integrato, fondato su una responsabilità condivisa tra istituzioni e cittadini e orientato alla costruzione di una condizione di "sicurezza cognitiva". Essa non si realizza perseguendo una linea difensiva, ma implementando altre misure, che favoriscano la co-regolamentazione delle piattaforme, la trasparenza sugli algoritmi e sui modelli di business che traggono profitto dalla diffusione di contenuti divisivi, la supervisione pubblica e istituzionale dei mezzi digitali. Soprattutto, essa riguarda la sfera informativa, la sfera culturale e quella educativa, dove gli obiettivi da perseguire sono principalmente l'innalzamento dei livelli d'istruzione, l'alfabetizzazione mediatica, il pluralismo delle voci pubbliche e il rafforzamento del pensiero critico. Si tratta, in estrema sintesi, di potenziare la cultura dei diritti e dell'inclusività e di dotare i cittadini di strumenti cognitivi e culturali per individuare e fare fronte a tutto ciò che può mettere in pericolo i principi democratici e i valori costituzionali.

Anche in relazione all'IA si evince che il problema della disinformazione non può essere risolto pensando solo a dover correggere la verità o falsità degli output. Come da più voci messo in rilievo, le informazioni sono condivise e diffuse perché capaci di rispondere alle più varie esigenze dell'utente. L'IA non crea questa dinamica, ma la rende più radicale, più stabile, più efficace. Nemmeno può di per sé aiutare a verificare fatti, a distinguere tra evidenze e apparenze. Essa genera testi coerenti, che, pur non avendo nulla a che vedere con la fattualità, producono una forte convinzione di accedere a una conoscenza ben fondata.

Poiché le democrazie si fondano sulla necessità che i cittadini condividano i significati necessari a muoversi nella realtà sociale, esse risultano più vulnerabili quando si radicalizzano fenomeni (culturali, politici, economici, o tecnologici) che colonizzano quei significati, incidendo sulla discussione collettiva e sulla percezione di ciò che è affidabile. In questo quadro, le democrazie sono più deboli se si trovano di fronte a un sistema dove, al pari di quanto accade nei sistemi totalitari, domina il principio dell'affabulazione narrativa, dell'incantamento collettivo.

È la ragione per cui, a tutela di una informazione che sostenga le democrazie, riteniamo che una grande attenzione debba essere profusa nella salvaguardia del sapere esperto e delle conoscenze storiche e scientifiche che l'umanità ha prodotto e vagliato fino ad ora, dei diritti conquistati.

Il disordine informativo è, oggi, ancora più chiaramente il condensato di un cassetto semantico di parole e dispositivi capaci di farci vedere la realtà nei termini di chi vi incide, di chi le inventa, di chi le diffonde. Sarebbe fuorviante non cogliere che tale cassetto è frutto di strategie raffinate e complesse che si avvalgono anche di quanto i processi di machine learning consentono di fare. Ma pensare unicamente in termini di condizioni tecnologiche è ancora più fuorviante. La tecnologia non opera in modo autonomo ma si iscrive nelle pratiche sociali e nei processi regolativi che ne orientano gli usi, gli sviluppi e gli effetti. Di fronte ai rischi della disinformazione, le democrazie possono trovare le risposte più efficaci per fronteggiarla nel responsabile e corretto funzionamento di tutte le istituzioni, soprattutto quelle deputate al bilanciamento e al controllo dei poteri, e nell'applicazione del principio di accountability da parte dei cittadini.

RIFERIMENTI BIBLIOGRAFICI

- Affuso O., Coco A. (2019), *Big Data: a challenge for social sciences. Criticisms for social knowledge and politics*, in «Central European Political Science Review», vol. 20, n. 76, pp. 93–109.
- Altheide D.L., Snow R.P. (1979), *Media Logic*, Beverly Hills (CA): Sage Publications.
- Antinori A. (2023), *Disinformazione e propaganda russa in Occidente*, in Casini E., Manciuoli A. (a cura di), *La guerra tiepida. Il conflitto ucraino e il futuro dei rapporti tra Russia e Occidente*, Roma: Luiss University Press, pp. 123–136.
- Bail C.A. et al. (2018), *Exposure to opposing views on social media can increase political polarization*, in «Proceedings of the National Academy of Sciences», vol. 115, pp. 9216–9221.
- Bender E.M. et al. (2021), *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, in «Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency», pp. 610–623.
- Bentivegna S., Boccia Artieri G. (2021), *Voci della democrazia. Il futuro del dibattito pubblico*, Bologna: il Mulino.
- Bontridder N., Pouillet Y. (2021), *The role of artificial intelligence in disinformation*, in «Data & Policy» 3: e32.
- Bonfanti M.E. (2021), *Impiego offensivo di contenuti digitali*, in Gori U., Vernon De Mars D. (a cura di), *Cyber Warfare 2019–2020. Dall'evoluzione della Warfare alla resilienza al Covid-19*, Milano: Franco Angeli, pp. 125–138.
- Cassidy J. (2021), *How post-truth politics transformed and shaped the outcome of the 2016 Brexit referendum*, in Piras S., Giusti E. (Eds.), *Democracy and Fake News. Information Manipulation and Post-Truth Politics*, London–New York: Routledge, pp. 53–63.
- Chadwick A. (2013), *The Hybrid Media System: Politics and Power*, Oxford, Oxford University Press.
- Cinelli M. et al. (2020), *The COVID-19 social media infodemic*, in «Scientific Reports», vol. 10, pp. 1–10.
- Comai G. (2021), *Responding to alleged Russian interference by focussing on the vulnerabilities that make it possible*, in Giusti S., Piras E. (Eds.), *Democracy and Fake News. Information Manipulation and Post-Truth Politics*, London–New York: Routledge, pp. 143–154.
- Costabile A. (2002), *Il potere politico*, Roma: Carocci.
- Cusumano E., Corbe M. (2018, eds.), *A Civil-Military Response to Hybrid Threats*, New York: Palgrave Macmillan.
- Del Vicario M. et al. (2016), *The spreading of misinformation online*, in «Proceedings of the National Academy of Sciences», vol. 113, pp. 554–559.
- Deutsch K.W. (1972), *I nervi del potere*, Milano: Etas Kompass.
- Di Pasquale M., Germani L.S. (2023), *Disinformaia e misure attive*, Research Paper, Istituto Gino Germani di Scienze Sociali e Studi Strategici.
- Edelman (2022), *Trust Barometer*, <https://www.edelman.com/trust/2022-trust-barometer>

- European Commission (2018), *Action Plan against Disinformation*, Joint Communication to EU Institutions, Brussels.
- European Commission (2020), *Security Union Strategy 2020–2025*, Brussels.
- Floridi L. (2017), *La quarta rivoluzione. Come l'infosfera sta cambiando il mondo*, Milano: Raffaello Cortina.
- Germani L.S. (2017), *La minaccia della disinformazione*, in Germani L.S. (a cura di), *Disinformazione e manipolazione delle percezioni*, Roma: Eurilink University Press.
- Giglietto F. et al. (2019), "Fake news" is the invention of a liar..., in «Current Sociology», vol. 67, n. 4, pp. 625–642.
- Giungato L., Taddei L., Affuso O., *Measuring Fabulation in Russo-Ukrainian War Online Narratives*, in «Italian Sociological Review», pp. 575–601.
- Giusti S., Piras E. (2021, eds), *Democracy and Fake News. Information Manipulation and Post-Truth Politics*, London–New York: Routledge.
- Gusfield J.R. (1984), *The Culture of Public Problems*, Chicago: University of Chicago Press.
- Habermas J. (2023), *Nuovo mutamento della sfera pubblica e politica deliberativa*, Milano: Raffaello Cortina.
- Hazelton A. (2021), *Once upon Covid-19...*, in Giusti S., Piras E. (Eds.), *Democracy and Fake News*, London–New York: Routledge, pp. 92–104.
- Jack C. (2017), *Lexicon of Lies: Terms for Problematic Information*, in «Data & Society», vol. 3, n. 22, pp. 1094–1096.
- Jungherr, A. (2023), *Artificial Intelligence and Democracy: A Conceptual Framework*, in «Social Media + Society», 9(3).
- Kertysova K. (2018), Artificial Intelligence and Disinformation, in «Security and Human Rights», 29 (1-4), pp. 55-81, https://brill.com/view/journals/shrs/29/1-4/article-p55_55.xml
- Lazer D.M. et al. (2018), *The science of fake news*, in «Science», vol. 359, pp. 1094–1096, 2018.
- Lorua E. et al. (2025), *The simulation of judgment in LLMs*, in «PNAS», vol. 122, n. 42, pp. 1–10.
- Lovari A., Ducci G. (2022), *Comunicazione pubblica. Istituzioni, pratiche, piattaforme*, Milano: Mondadori.
- Luhmann N. (2002), *La fiducia*, Bologna: il Mulino.
- LUISS (2023), *Come individuare e contrastare operazioni coordinate di disinformazione in Italia. Casi di studio e indicazioni di policy*, Roma.
- Mutti A. (1998), *Capitale sociale e sviluppo. La fiducia come risorsa*, Bologna: il Mulino.
- NATO (2014), *Hybrid War – Hybrid Response?*, in «NATO Review Magazine».
- NATO (2021), *NATO 2030: United for a New Era*, Brussels.
- Quattrociocchi W., Vicini A. (2023), *Polarizzazioni. Informazioni, opinioni e altri demoni dell'infosfera*, Milano: Franco Angeli.

OECD (2020), *Governance Responses to Disinformation: How Open Government Principles Can Inform Policy Options*, Paris.

OECD (2021), *OECD Report on Public Communication. The Global Context and the Way Forward*, Paris, OECD Publishing.

Panbianco A. (1982), *Modelli di partito*, Bologna: il Mulino.

Presidenza del Consiglio dei Ministri (2022), *Relazione sulla politica dell'informazione per la sicurezza*.

Simmel G. (1984), *Filosofia del denaro*, Torino: Utet.

Zafesova A. (2021), *Lie to Live...*, in Giusti S., Piras E. (Eds.), *Democracy and Fake News*, London–New York: Routledge.